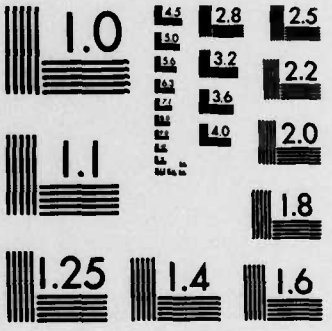
AD-A132 328 SSING SIMULATION TO ESTIMATE FIRST PASSAGE DISTRIBUTIONS(U) CALIFORNIA UNIV BERKELEY OPERATIONS RESEARCH CENTER S M ROSS ET AL. JAN 83 ORC-83-1 1/1

UNCLASSIFIED AFOSR-81-0122 F/G 12/1 NL



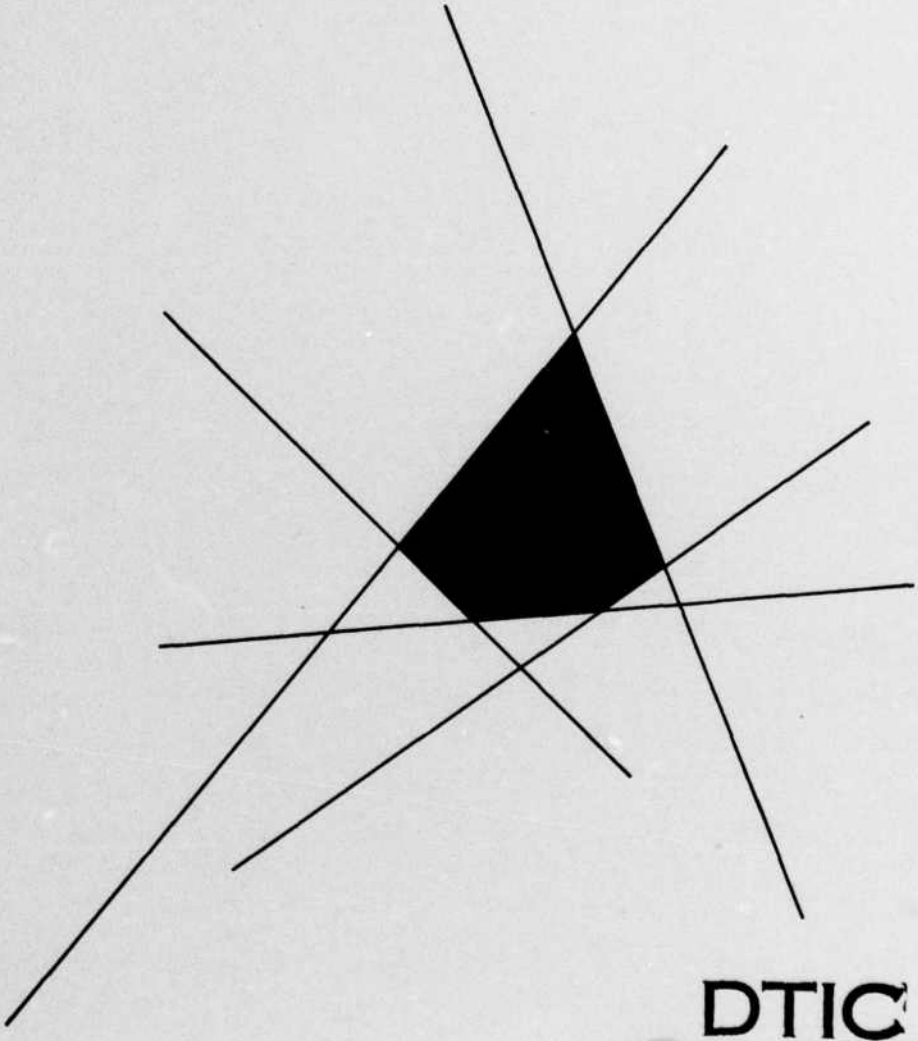
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

ORC 83-1
JANUARY 1983

12

USING SIMULATION TO ESTIMATE FIRST PASSAGE DISTRIBUTIONS

by
SHELDON M. ROSS
and
ZVI SCHECHNER



OPERATIONS RESEARCH CENTER

UNIVERSITY OF CALIFORNIA · BERKELEY

USING SIMULATION TO ESTIMATE FIRST PASSAGE DISTRIBUTIONS

by

Sheldon M. Ross
Department of Industrial Engineering
and Operations Research
University of California, Berkeley

and

Zvi Schechner
Department of Industrial Engineering
and Operations Research
Columbia University
New York, New York

JANUARY 1983 ORC 83-1

This research was supported by the Air Force Office of Scientific Research (AFSC), USAF, under Grant AFOSR-81-0122 with the University of California. Reproduction in whole or in part is permitted for any purpose of the United States Government.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER: ORC 83-1	2. GOVT ACCESSION NO.: AD-A132328
	3. RECIPIENT'S CATALOG NUMBER:
4. TITLE (and Subtitle): USING SIMULATION TO ESTIMATE FIRST PASSAGE DISTRIBUTIONS	5. TYPE OF REPORT & PERIOD COVERED: Research Report
	6. PERFORMING ORG. REPORT NUMBER:
7. AUTHOR(s): Sheldon M. Ross and Zvi Schechner	8. CONTRACT OR GRANT NUMBER(s): AFOSR-81-0122
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Operations Research Center, University of California, Berkeley, California 94720	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 2304/A5
11. CONTROLLING OFFICE NAME AND ADDRESS: United States Air Force, Air Force Office of Scientific Research, Bolling Air Force Base, D.C. 20332	12. REPORT DATE: January 1983
	13. NUMBER OF PAGES: 24
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report): Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Simulation
First Passage
Discrete Hazard
Discrete Time
Continuous Time
Convolution
Delta Method
Uniformizing

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

(SEE ABSTRACT)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-LF-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

ABSTRACT

Consider a discrete time Markov process $\{X_n, n \geq 0\}$. For a given subset A of the state space consider the problem of using simulation to estimate the number of transitions it takes the process to enter A. Using estimators based on the "observed hazard", we are able to improve on the usual Monte Carlo estimator. We also consider the problem of estimating the distribution of the first state in A to be reached, and then extend our results to continuous time.

USING SIMULATION TO ESTIMATE FIRST PASSAGE DISTRIBUTIONS

by

Sheldon M. Ross and Zvi Schechner

0. INTRODUCTION

Consider a discrete time Markov process $\{X_n, n = 0,1, \ldots\}$ such that whenever the present state is x the next state is chosen according to the distribution P_x. Let $X_0 = 0$ be fixed and consider for a given set of states A the number of transitions N until the Markov process enters the set A. We are interested in estimating the distribution and the mean of N by use of simulation.

Assuming that the process $\{X_n\}$ is such that N is finite with Probability 1, this model can be simulated by the standard Monte Carlo method of letting $X_0 = 0$ and then generating a random variable from the distribution P_0 to determine X_1. If $X_1 = x$, we then generate a random variable from the distribution P_x and set it equal to X_2, and so on. We stop the run when we obtain a state in A. The process is then repeated for a second run, and so on until a total of r (a fixed predetermined number) simulation runs are obtained. If we let $N^{(i)}$, $i = 1, \ldots, r$ denote the number of steps until A is reached in the ith run, then the usual way to estimate $P(N = k)$, $k \geq 1$ and $E[N]$ is as follows:

$$P(N = k) \text{ is estimated by } \#\{i : N^{(i)} = k\}/r$$

$$E[N] \text{ is estimated by } \sum_{i=1}^{r} N^{(i)}/r .$$

In this paper, we propose new estimators for these quantities which are based on the "observed hazards" also known as the "predictable projection".

That is, consider a given simulation run terminating at N and define

$$\lambda_n = P_{X_{n-1}}(A) \quad \text{if} \quad N \geq n$$

where X_{n-1} is the $(n-1)$th state and $P_x(A)$ is the probability that the next state visited from x is in A. We use the process $\{\lambda_n\}$ to estimate the various quantities of interest. In Section 1, we present two estimators for $E[N]$. The first of these can be used to form an interval estimate and we show that it is unbiased, having a smaller variance than the classical Monte Carlo estimator. The second estimator presented is a point estimate of $E[N]$. In Section 2, we consider the problem of estimating the distribution of N. In Section 3, we consider the distribution of the final state X_N. Section 4 deals with the continuous time analog and as a consequence a new formula for estimating convolution is obtained.

1. ESTIMATING THE MEAN NUMBER OF TRANSITIONS

1.1 Method 1

For a given simulation run if we let N denote the number of transitions needed to enter A, then N can be expressed as

$$N = \sum_{j=1}^{\infty} I_j$$

where

$$I_j = \begin{cases} 1 & \text{if } j \leq N \\ 0 & \text{if } j > N . \end{cases}$$

Hence,

$$E[N] = \sum_{j=1}^{\infty} E[I_j] = 1 + \sum_{j=2}^{\infty} E[I_j]$$

and the usual estimator can be regarded as estimating $E[I_j]$ by looking at the values of I_j for that cycle and then summing over j to estimate $E[N]$.

We propose to estimate I_j by conditioning on the first $j - 2$ transitions. As

$$E[I_j \mid X_0, X_1, \ldots, X_{j-2}] = P\{N \geq j \mid X_0, \ldots, X_{j-2}\}$$

$$= \begin{cases} 1 - P_{X_{j-2}}(A) & \text{if } N > j - 2 \\ 0 & \text{if } N \leq j - 2 . \end{cases}$$

Hence, our estimate for $E[N]$ based on a single simulation run ending when the process enters A is

$$1 + \sum_{j=2}^{N+1} \left[1 - P_{X_{j-2}}(A)\right] = N + 1 - \sum_{j=2}^{N+1} P_{X_{j-2}}(A)$$

$$= N + 1 - \sum_{j=1}^{N} \lambda_j \ .$$

If r simulation runs are made, we propose to estimate $E[N]$ by the average of the above single run estimators.

As

$$E[E[I_j \mid X_0, \ldots, X_{j-2}]] = E[I_j] \ ,$$

it follows that our estimator is unbiased. Also by the conditional variance formula

$$\text{Var}\ (E[I_j \mid X_0, \ldots, X_{j-2}]) \leq \text{Var}\ (I_j) \ .$$

However, the above does not imply that

$$\text{Var}\left(\sum_{j=2}^{\infty} E[I_j \mid X_0, \ldots, X_{j-2}]\right) \leq \text{Var}\left(\sum_{j=2}^{\infty} I_j\right) \ .$$

We now prove the above inequality, thus showing that our estimate has smaller variance than N .

Proposition 1:

$$\text{Var}\left(\sum_{j=2}^{\infty} E[I_j \mid X_0, \ldots, X_{j-2}]\right) \leq \text{Var}\ [N] \ .$$

Proof:

Let $H_i = \{X_0, \ldots, X_i\}$ denote the history up to time i and define

$$\hat{I}_i = E[I_i \mid H_{i-2}] \ .$$

Now,

$$\text{Var}\left(\sum_j I_j\right) = \sum_j \text{Var}\ (I_j) + \sum_{i \neq j}\sum \text{Cov}\ (I_i, I_j)$$

and

$$\text{Var}\left(\sum_j \hat{I}_j\right) = \sum_j \text{Var}\ (\hat{I}_j) + \sum_{i \neq j}\sum \text{Cov}\ (\hat{I}_i, \hat{I}_j)\ .$$

As $\text{Var}\ (\hat{I}_j) \leq \text{Var}\ (I_j)$ and $E[\hat{I}_j] = E[I_j]$, it suffices to prove that

$$E[\hat{I}_i \hat{I}_j] \leq E[I_i I_j] \quad \text{for all} \quad i \neq j\ .$$

However, for $i < j$,

$$\begin{aligned} E[\hat{I}_i \hat{I}_j] &= E[E[I_i \mid H_{i-2}]E[I_j \mid H_{j-2}]] \\ &= E[E[\hat{I}_i I_j \mid H_{j-2}]] \\ &= E[\hat{I}_i I_j] \\ &= E[I_j] \qquad \text{since} \quad 0 \leq \hat{I}_i \leq 1 \\ &= E[I_i I_j] \qquad \text{since} \quad I_i I_j = I_j \end{aligned}$$

which proves the result.

Remark:

It is illuminating to write the proposed estimator of $E(N)$ as

$$N + \left(1 - \sum_{j=1}^{N} \lambda_j\right)\ .$$

As $\sum_{j=1}^{N} \lambda_j$, the total hazard, has mean 1, the above is (as we know) unbiased. Intuitively, it has smaller variance than N for in simulation runs for which N is small (large) the total hazard will also (usually) tend to be small (large) and so the additional term in parenthesis acts as a correction factor.

Of course, an even better estimator is

$$N + \alpha\left(1 - \sum_{j=1}^{N} \lambda_j\right)$$

where $\alpha = \dfrac{\text{Cov}\left(N, \sum_{j=1}^{N} \lambda_j\right)}{\text{Var}\left(\sum_{1}^{N} \lambda_j\right)}$, but the difficulty with this is, of course, that α is usually not known.

1.2 Method 2

Suppose that we make a total of r simulation runs. For run i of length $N^{(i)}$, define for $n \leq N^{(i)}$

$$\lambda_n^i = P_{X_{n-1}^i} \quad (A)$$

where X_{n-1}^i is the $(n-1)$st state visited in the ith run. Now, for a given n , consider each run in which $N^{(i)} \geq n$ and let

$$\bar{\lambda}_n = \sum_{i:N^{(i)} \geq n} \lambda_n^i \Big/ r(n)$$

where $r(n)$ = number of i : $N^{(i)} \geq n$. Hence, $\bar{\lambda}_n$ is an estimate of the discrete hazard (or failure) rate value $P\{N = n \mid N \geq n\}$. As

$$E[N] = \sum_{j=2}^{\infty} P\{N \geq j\} + 1$$

$$= \sum_{j=2}^{\infty} \prod_{n=1}^{j-1} P\{N > n \mid N \geq n\} + 1 ,$$

we have as a point estimate of $E[N]$

$$\sum_{j=2}^{\infty} \prod_{n=1}^{j-1} (1 - \bar{\lambda}_n) + 1$$

where $\bar{\lambda}_n$ is taken to equal 1 for $n > \max_i N^{(i)}$.

1.3 Numerical Example

Let $\{X_n , n \geq 0\}$ be a Markov chain having the following transition matrix:

$$\begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \begin{pmatrix} .23 & .5 & .25 & .02 \\ .2 & .4867 & .3 & .0133 \\ .6 & .2 & .19 & .01 \\ 0 & 0 & 0 & 1 \end{pmatrix} \end{array}$$

Letting $X_0 = 0$, the process $\{X_n\}$ has been simulated 1000 times to determine the expected first passage time to state 4.

	E[N]	standard deviation of estimate
theoretical	68.281	--
crude	66.988	2.139
method 1	67.015	2.108
method 2	68.059	?

2. ESTIMATION OF $P\{N > j\}$

2.1 Method 1

As shown in Section 1.2, the quantity $\bar{\lambda}_n$ is an estimate of $P\{N = n \mid N \geq n\}$, and thus $\prod_{n=1}^{j} (1 - \bar{\lambda}_n)$ is an estimator of $P\{N > j\}$. To approximate the variance of this estimator, we will make use of the so-called Delta Method used to approximate the mean and variance of a function $g(X)$ when $E[X] = \mu$, $\text{Var}(X) = \sigma^2$ are known. This method approximates $g(X)$ by

$$g(X) \approx g(\mu) + g'(\mu)(X - \mu)$$

leading to

$$E(g(X)) \approx g(\mu)$$

$$\text{Var}(g(X)) \approx (g'(\mu))^2 \sigma^2 .$$

We now show how to approximate $\text{Var}\left(\prod_{n=1}^{j} (1 - \bar{\lambda}_n)\right)$.

Let

$$\bar{P}_n = 1 - \bar{\lambda}_n \quad , \quad M_k = \prod_{n=1}^{k} \bar{P}_n .$$

That is, to obtain $\bar{P}_n$, one considers each of the $r(n)$ simulation runs which do not end before n transitions and for the runs it takes the average probability that the nth transition does not take the process into A. Hence, when defined, $\bar{P}_n$ is an unbiased estimate of $P_n \equiv P\{N > n \mid N \geq n\}$.

Now,

$$\log M_k = \sum_{n=1}^{k} \log \bar{P}_n$$

and so

$$\text{Var } (\log M_k) \approx \sum_{n=1}^{k} \text{Var } (\log \bar{P}_n)$$

$$\approx \sum_{n=1}^{k} \text{Var } (\bar{P}_n)/P_n^2$$

where the first approximation results from assuming that the $\bar{P}_n$ are independent and the second from applying the Delta Method. As

$$\text{Var } (\bar{P}_n) = \text{Var } (1 - \bar{P}_n)$$

$$= \left[E\left[P^2_{X_{n-1}}(A) \mid N \geq n - 1 \right] - q_n^2 \right] / r(n)$$

where $q_n = 1 - P_n = P\{N = n \mid N \geq n\}$, we have

$$\text{Var } (\log M_k) \approx \sum_{n=1}^{k} \frac{\left(E\left[P^2_{X_{n-1}}(A) \mid N \geq n - 1 \right] - q_n^2 \right)}{P_n^2 r(n)} .$$

Again, using the Delta Method, we have

$$\text{Var } (\log M_k) \approx \text{Var } (M_k)/E^2[M_k]$$

and so

$$\text{Var}\ (M_k) \approx E^2[M_k] \sum_{n=1}^{k} \frac{E\left[P^2_{X_{n-1}}(A) \mid N \geq n-1\right] - q_n^2}{r(n)P_n^2} .$$

Hence, we can obtain a rough estimate of the variance of our estimate M_j by using the above and approximating $E[M_j]$, $E\left[P^2_{X_{n-1}}(A) \mid N \geq n-1\right]$, P_n, q_n by

$$E[M_j] \overset{\text{est}}{=} M_j$$

$$E\left[P^2_{X_{n-1}}(A) \mid N \geq n-1\right] \overset{\text{est}}{=} \sum_{i:N_i \geq n} \left(P_{X^i_{n-1}}(A)\right)^2 \Big/ r(n)$$

$$P_n \overset{\text{est}}{=} \sum_{i:N_i \geq n} P_{X^i_{n-1}}(A) \Big/ r(n)$$

$$q_n \overset{\text{est}}{=} 1 - P_n$$

where $\overset{\text{est}}{=}$ means "is estimated by".

2.2 Method 2

A second method for estimating $P(N \leq j)$ is

$$\frac{1}{r} \sum_{i=1}^{r} A_j^{(i)}$$

where $A_j^{(i)} = \sum_{n=1}^{j} \lambda_n^{(i)}$, and where $\lambda_n^{(i)} = 0$ if $N^{(i)} < n$.

It is unbiased and since the different simulation runs are independent, its variance is $\frac{1}{r}\,\text{Var}\left(A_j^{(1)}\right)$.

Proposition:

$$\text{Var}\left(A_j^{(1)}\right) = 2E\left[A_j^{(1)} 1\{N^{(1)} \le j\}\right] - E\left[\lambda_{N^{(1)}} 1\{N^{(1)} \le j\}\right] - \{P(N^{(1)} \le j)\}^2$$

where $1\{\cdot\}$ is the indicator function of the specified event $\{\cdot\}$.

Proof:

$$\left(A_j^{(1)}\right)^2 = \left(\sum_{n=1}^{j} \lambda_n^{(1)}\right)^2$$

$$= 2 \sum_{k=1}^{j} \sum_{n=1}^{k} \lambda_n^{(1)} \lambda_k^{(1)} - \sum_{n=1}^{j} \left(\lambda_n^{(1)}\right)^2$$

$$= 2 \sum_{n=1}^{j} A_n^{(1)} \lambda_n^{(1)} - \sum_{n=1}^{j} \left(\lambda_n^{(1)}\right)^2 .$$

Taking expectation of both sides, we obtain

$$E\left[\left(A_j^{(1)}\right)^2\right] = 2 \sum_{n=1}^{j} E\left[A_n^{(1)} \lambda_n^{(1)}\right] - \sum_{n=1}^{j} E\left[\left(\lambda_n^{(1)}\right)^2\right] .$$

Using the basic property of conditional expectation, we get

$$E\left[\left(A_j^{(1)}\right)^2\right] = 2 \sum_{n=1}^{j} E\left[A_n^{(1)} 1\{N^{(1)} = n\}\right] - \sum_{n=1}^{j} E\left[\lambda_n^{(1)} 1\{N^{(1)} = n\}\right]$$

$$= 2E\left[A_{N^{(1)}}^{(1)} 1\{N^{(1)} \le j\}\right] - E\left[\lambda_{N^{(1)}}^{(1)} 1\{N^{(1)} \le j\}\right]$$

and the result follows.

The above proposition provides good insight to the condition under which the proposed method is superior to the classical crude method. Consider a chain in which the set A is "rare", that is, for $x \notin A$, $p_1 \leq P_x(A) \leq p_2$ where $0 < p_1 < p_2$ are "small", then for a moderate j, $P(N^{(1)} \leq j) = p$ is small and the variance which appears in the crude method is (roughly) proportional to p/r whereas the variance of the proposed estimator is of the order of magnitude of p^2/r. For small p, this is a significant improvement.

2.3 Numerical Example

$\{X_n, n \geq 0\}$ is as in 1.3. It was simulated 1000 to determine P(enters state 3 by time 10).

Probability and Standard Deviation of Estimate

	Probability	Standard Deviation
theoretical	.1408	--
crude	.1480	1123×10^{-5}
method 1	.1407	5.5×10^{-5}
method 2	.1408	93.5×10^{-5}

Relative efficiencies:

method 1 to crude: 205
method 2 to crude: 12

3. ESTIMATION OF $P\{X_N \in B\}$

Let $B \subset A$ denote a subset of A and consider the problem of estimating the probability that the final state is in B. To do this, let N_B equal the number of transitions until B is reached (and so $N_B = \infty$ if $X_N \notin B$).

Our approach to estimating $P\{X_N \in B\}$ will consist of working with the following modified version of the hazard rate function of N_B--namely,

$$\lambda_B^*(n) = P\{N_B = n \mid N \geq n\} .$$

To estimate $\lambda_B^*(n)$, consider those simulation runs for which $N^{(i)} > n - 1$ and then take as the estimate of $\lambda_B^*(n)$ the average value of $P_{X_{n-1}}(B)$ for those runs. That is, $\bar{\lambda}_B(n)$, the estimate of $\lambda_B^*(n)$ is given by

$$\lambda_B^*(n) = \sum_{i:N^{(i)} \geq n} P_{X_{n-1}^i}(B) \Big/ r(n)$$

where $r(n)$ is the number of runs for which $N \geq n$. Take $\lambda_B^*(n)$ to equal 0 if $r(n) = 0$.

Hence, we have an estimate for $P\{N_B = n\}/P\{N \geq n\}$. Also, by using the method of Section 2.1, we can estimate $P\{N \geq n\}$ and so by using the ratio of these estimators we have a method for estimating $P\{N_B = n\}$. By summing over all n, we can then estimate $\sum_{n=1}^{\infty} P\{N_B = n\} = P\{X_N \in B\}$. In addition, as we are assuming, but not using in our estimation procedure, the result that $P\{X_N \in A\} = 1$, we can improve our estimate of $P\{X_N \in B\}$--call it $\hat{P}\{X_N \in B\}$--by normalizing. That is, our estimate of $P\{X_N \in B\}$ is $\hat{P}\{X_N \in B\}/\hat{P}\{X_N \in A\}$.

Remark:

The reason we use the modified hazard rate function $P\{N_B = n \mid N \geq n\}$ instead of the usual hazard rate function $P\{N_B = n \mid N_B \geq n\}$ is that a direct estimate of this latter expression does not use all available information. For if $N_B \geq n$ then $P_{X_{n-1}}(B)$ estimates $P\{N_B = n \mid N_B \geq n\}$, but in addition $P_{X_{n-1}}(A - B)$ yields valuable information about $P\{X_N \notin B\}$. This information is captured in our method by coming into play in our estimate of $P\{N \geq n\}$. Hence, our ratio estimate makes use at each stage of both the probability that the next state will be in B and the probability that B will never be reached. A product form estimate of $P\{N_B < \infty\}$ which uses only direct estimates of the (unmodified) hazard rate function will not utilize this latter information.

4. CONTINUOUS TIME VERSION

Suppose now that the Markov process is in continuous time-spending and exponential amount of time with mean $\mu(x)$, $0 < \mu(x) < \infty$, when in state x before transiting into another state according to P_x . Suppose also that the Markov chain is regular in the sense that with Probability 1 it only makes a finite number of transitions in any finite time interval. Let T denote the time it takes for the process to enter A .

4.1 Estimating E[T]

In analogy with Proposition 1, for any given simulation run, an estimate with smaller variance and the same mean as T is

$$T + 1 - \text{total hazard} = T + 1 - \int_0^T P_{X(s)}(A)/\mu(X(s))ds$$

where $X(s)$ is the state at time s . Now, let $X_0, X_1, \ldots, X_N$ denote the sequence of states visited. The above estimator can be improved thusly:

$$E[T + 1 - \text{total hazard} \mid N, X_0, X_1, \ldots, X_N]$$

$$= E[T \mid N, X_0, \ldots, X_N] + 1 - E[\text{hazard} \mid N, X_0, \ldots, X_N]$$

$$= \mu(X_0) + \ldots + \mu(X_{N-1}) + 1 - E\left[\sum_{i=0}^{N-1} \tau_i P_{X_i}(A)/\mu(X_i)\right]$$

$$= \mu(X_0) + \ldots + \mu(X_{N-1}) + 1 - \sum_{i=0}^{N-1} P_{X_i}(A)$$

where τ_i in the above represented the time spent in state X_i before the transition to X_{i+1} occurs (and is thus exponential with mean $\mu(X_i)$ independent of the next state visited).

Hence, our estimate for $E[T]$ in any run is just the conditional value of T given the sequence of states visited plus 1 minus the total discrete hazard given these states. The overall estimate is then obtained by averaging this quantity over all simulation runs.

4.2 Estimating $P\{T \leq t\}$

Again we recommend conditioning on the sequence of states visited $N, X_0, X_1, \ldots, X_N$. That is, for a given simulation run, let

$$I = \begin{cases} 1 & \text{if } T \leq t \\ 0 & \text{otherwise .} \end{cases}$$

Then,

$$E[I \mid N, X_0, \ldots, X_N] = P\{\tau_0 + \ldots + \tau_{N-1} \leq t\}$$

where $\tau_0, \ldots, \tau_{N-1}$ are independent exponentials with respective means $\mu(X_0), \mu(X_1), \ldots, \mu(X_{N-1})$. Where appropriate ($N$ relatively large and none of the $\mu(X_0), \ldots, \mu(X_{N-1})$ dominant) the above can be approximated by the central limit theorem and where this is not appropriate a simulation can be done using the approach of Section 4.2.1.

4.2.1 Simulation Estimation of a Convolution

We start with a proposition which will be necessary in the sequel.

Proposition:

Let $\underline{X} = (X_1, X_2, \ldots)$ be any sequence of random variables and let $g_i(\underline{X})$ denote arbitrary functions, $i = 1, \ldots, n$. Then, for $\lambda_i \geq 0$, $\sum_{i=1}^{n} \lambda_i = 1$, we have that for any random variable Z

$$\text{Var}\left(\sum_{i=1}^{n} \lambda_i E[Z \mid g_i(\underline{X})]\right) \leq \text{Var}\ (Z)\ .$$

Proof:

Let M denote a random variable independent of Z and of $\underline{X}$ and such that

$$P\{M = i\} = \lambda_i\ ,\quad i = 1, \ldots, n\ .$$

Then, by the conditional variance formula,

$$\begin{aligned}
\text{Var}\ (Z) &\geq \text{Var}\ (E[Z \mid M, g_M(\underline{X})]) \\
&= E[\text{Var}\ (E[Z \mid M, g_M(\underline{X})] \mid \underline{X})] \\
&\quad + \text{Var}\ (E[E[Z \mid M, g_M(\underline{X})] \mid \underline{X}]) \\
&\geq \text{Var}\ (E[E[Z \mid M, g_M(\underline{X})] \mid \underline{X}]) \\
&= \text{Var}\left(\sum_{i=1}^{n} \lambda_i E[Z \mid g_i(\underline{X})]\right)\ .\ ||
\end{aligned}$$

Now, let $X_1, X_2, \ldots, X_n$ be independent with X_i having distribution F_i and suppose we want to use simulation to estimate $P\{X_1 + \ldots + X_n \leq t\}$. To do so, let

$$Z = I = \begin{cases} 1 & \text{if } \sum_{i=1}^{n} X_i \leq t \\ 0 & \text{otherwise}\ . \end{cases}$$

We can use the above by letting

$$g_i(\underline{X}) = \sum_{j \neq i} X_j$$

and so

$$E[I \mid g_i(\underline{X})] = F_i\left(t - \sum_{j \neq i} X_j\right)$$

and so an unbiased estimate of the convolution having smaller variance than I is given by

$$P\{X_1 + \ldots + X_n \leq t\} \overset{\text{est}}{=} \sum_{i=1}^{n} \lambda_i F_i\left(t - \sum_{j \neq i} X_j\right)$$

where $\lambda_i \geq 0$, $\sum_{i=1}^{n} \lambda_i = 1$.

As the above estimator is an improvement over I for all nonnegative vectors $\underline{\lambda}$ that sum to unity, it remains to present a method for picking the "optimal" $\underline{\lambda}$. To do so, let us recall that if $W_1, W_2, \ldots, W_n$ are independent random variables having common mean then variance of $\sum_{i=1}^{n} \lambda_i W_i$ is minimized by letting $\lambda_i = \dfrac{1/\text{Var}\ (W_i)}{\sum_{j=1}^{n} 1/\text{Var}\ (W_j)}$. Whereas $F_i\left(t - \sum_{j \neq i} X_j\right)$, $i = 1, \ldots, n$ are clearly not independent, let us nevertheless use this result to choose the convex linear coefficients $\underline{\lambda}$. Hence, we must determine $\text{Var}\ F_i\left(t - \sum_{j \neq i} X_j\right)$.

Now, if Y is a random variable and if F is a distribution function with density $F' = f$, then by the Delta Method

$$\text{Var}\ (F(t - Y)) \approx \text{Var}\ (Y)[f(t - E[Y])]^2 .$$

Hence,

$$\text{Var}\left[F_i\left(t - \sum_{j\neq i} X_j\right)\right] \approx \left[f_i\left(t - \sum_{j\neq i} \mu_j\right)\right]^2 \sum_{j\neq i} \sigma_j^2$$

where μ_j and σ_j^2 are respectively the mean and variance of F_j. We thus recommend taking

$$\lambda_i = \frac{\left(f_i^2\left(t - \sum_{j\neq i} \mu_j\right) \sum_{j\neq i} \sigma_j^2\right)^{-1}}{\sum_{i=1}^{n} \left(f_i^2\left(t - \sum_{j\neq i} \mu_j\right) \sum_{j\neq i} \sigma_j^2\right)^{-1}} .$$

In the special case where F_j is exponential with mean μ_j and so $f_j(x) = \frac{1}{\mu_j} e^{-x/\mu_j}$, $\sigma_j^2 = \mu_j^2$, we have

$$\lambda_i = C\left(\frac{e^{-2\left(t - \sum_{j\neq i} \mu_j\right)}}{\mu_i^2} \sum_{j\neq i} \mu_j^2\right)^{-1}$$

where C is chosen to make $\sum_{i=1}^{n} \lambda_i = 1$.

4.2.2 Estimating $P\{T \leq t\}$ by Uniformizing

Another possibility for estimating $P\{T \leq t\}$ is to uniformize the Markov process--which can be done assuming that $\inf_x \mu(x) > 0$. This transforms the Markov process into one which spends an exponential time with rate $\mu(x) \equiv \mu \equiv 1/\lambda$ in state x. (See, for instance, Ross [3]). Hence, we can compute $P\{T \leq t\}$ by conditioning on the number of transitions it takes to enter A. This gives

$$P\{T \leq t\} = \sum_n P\{T \leq t \mid N = n\}P\{N = n\}$$

$$= \sum_n G_{n,\lambda}(t)P\{N = n\}$$

where

$$G_{n,\lambda}(t) = \int_0^t \lambda e^{-\lambda x} \frac{(\lambda x)^{n-1}}{(n-1)!}\, dx = 1 - e^{-\lambda t} \sum_{j=0}^{n-1} \frac{(\lambda t)^j}{j!}$$

is the gamma (n,λ) distribution function. Hence, we can estimate $P\{T \leq t\}$ by uniformizing the process and then use the resultant simulation runs to estimate $P\{N = n\}$. This last quantity can be estimated either by

$$P\{N = n\} \overset{est}{=} \begin{cases} P_{X_{n-1}}(A) & \text{if } N \geq n \\ 0 & \text{if } N < n \end{cases}$$

and then averaging over all runs, or by using

$$P\{N = n\} \overset{est}{=} \bar{\lambda}_n \prod_{i=1}^{n-1} (1 - \bar{\lambda}_i) .$$

Remark:

By "uniformizing", we eliminate all variance due to the times spent in each state. However, we pay the price of increasing the variance on the number of transitions involved. This procedure is thus only recommended when the values $\mu(x)$ are relatively constant--for in this case very few additional transitions result.

REFERENCES

[1] Carter, G. and E.J. Ignall, "Virtual Measures: A Variance Reduction Technique for Simulation," Management Science, Vol. 21, No. 6, pp. 607-616, (1975).

[2] Hordijk, A., D. Iglehart and R. Schassberger, "Discrete Time Methods for Simulating Continuous Time Markov Chains," Advances in Applied Probability, Vol. 8, pp. 772-788, (1976).

[3] Ross, S.M., STOCHASTIC PROCESSES, John Wiley & Sons, New York, (1982).

END

FILMED

9-83

DTIC

$$\text{var}\,(\log \pi_k) \sim \text{var}\,(\pi_k)/E\,[\pi_k]^2$$

and so

It is unbiased and since the different aimulation runs are independent, its variance is $\frac{1}{r} \text{Var}\left(A_j^{(1)}\right)$.

and the result follows.